Санкт-Петербургский государственный университет

Математическое обеспечение и администрирование информационных систем

Технология программирования

Хлобыстова Анастасия Олеговна

Агрегирование данных из социальных сетей в целях упреждающей диагностики уязвимостей пользователей к социоинженерным атакам (проектная работа)

Бакалаврская работа

Научный руководитель:

к. т. н., ст. преп. М.В. Абрамов

Рецензент:

к. т. н., асс. И.Б. Сметанников

Санкт-Петербург

2019

SAINT-PETERSBURG STATE UNIVERSITY

Software and Administration of Information Systems

Technology of Programming

Khlobystova Anastasiia

Data aggregation from social networks for proactively diagnostics of user vulnerabilities to social engineering attacks (joint project)

Bachelor's Thesis

Scientific supervisor:

PhD Sci. (Eng.), Senior Lecturer Maksim Abramov

Reviewer:

PhD Sci. (Eng.), ass. Ivan Smetannikov

Saint-Petersburg
2019

Оглавление

Введение

Актуальность темы. Несмотря на рост эффективности и повышение качества средств защиты конфиденциальной информации от программно-технических атак, информационные системы остаются уязвимыми [47, 53]. Часто ключевую роль в инцидентах нарушения безопасности информации играет человек — санкционированный пользователь системы [26, 27, 31, 40, 43]. Атаки на пользователей информационных систем с использованием методов социальной инженерии стали происходить чаще, приносить большие убытки и требовать больше времени для расследования подобных преступлений [28]. Под социоинженерной атакой согласно [38] понимается набор прикладных психологических и аналитических приемов, которые злоумышленники применяют для скрытой мотивации пользователей публичной или корпоративной сети к нарушениям устоявшихся правил и политик в области информационной безопасности. Одним из подтверждений актуальности исследуемой тематики может служить информация, предоставляемая крупными российскими компаниями, свидетельствующая о применении методов социальной инженерии в 80% случаев от числа всех инцидентов нарушения безопасности [48]. Данная информация также отражена и в отчётах компаний, специализирующихся на разработке систем защиты от кибератак [40, 44]. Актуальность данного направления находит подтверждение и у экспертов в области информационной безопасности [32]. Согласно их прогнозам [54] в ближайшие несколько лет ожидается рост активности злоумышленников-социоинженеров.

Помимо сотрудников организаций жертвами социоинженерных атак также становятся и частные лица. Последние годы в России в этой роли оказывались 1,2 млн человек ежегодно [45]. В первом полугодии 2018 года у жертв в среднем было похищено 5000 рублей, что составляет 12% среднемесячной зарплаты россиянина [45]. По классификации,

представленной в монографии [39], особую группу социоинженерных атак составляют фишинг-атаки, кроме того, отмечается их существенный рост за последнее время. Данный факт также находит подтверждение в отчёте компании Group-IB [53]: при помощи веб-фишинга хакерам удалось украсть около $4,2 млн. При этом в среднем за день совершается более 1,2 тысяч социоинженерных атак [53].

В связи с этим актуальной видится задача защиты пользователей информационных систем от социоинженерных атак. Одной из мер, способствующих обеспечению защиты пользователей информационных систем от социоинженерных атак, является анализ защищенности. Общая цель направления исследований заключается в повышении уровня информационной безопасности организации за счет разработки автоматизированных средств анализа защищённости пользователей информационных систем от социоинженерных атак.

Часто социоинженерные атаки осуществляются через цепочку пользователей. Такие атаки называются многоходовыми социоинженерными атаками [2]. Атаковать целевого пользователя при этом можно через разные цепочки, и оценки вероятности прохождения по ним будут отличаться. Если представить сотрудников организации в виде социального графа, то можно говорить о разных траекториях реализации многоходовых социоинженерных атак. При этом оценки вероятности успешного прохождения этих траекторий отличаются. В связи с чем, возникает необходимость выявления наиболее вероятной траектории распространения многоходовой социоинженерной атаки или совокупности таких траекторий. Кроме того, важно учитывать, что от реализации разных траекторий организация несет отличающиеся по размеру убытки. Таким образом, существенно выявлять не только наиболее вероятные траектории, но разработать подход к идентификации наиболее критичных траекторий.

Степень разработанности темы. Исследования, направленные на изучение характера взаимодействия пользователей, квантификацию характеристик данного взаимодействия и его влияния на распространение социоинженерной атаки, проводились и продолжают проводиться на базе лаборатории теоретических и междисциплинарных проблем информатики Санкт-Петербургского института информатики и автоматизации РАН (ТиМПИ СПИИРАН). А именно, был разработан набор моделей: «критичные документы – информационная система – персонал – злоумышленник» [5]. А также исследованы вопросы по построению и анализу социальных графов сотрудников компании [2]. Предложены подходы идентификации аккаунтов сотрудников компании в социальных сетях [30]. Рассмотрены вопросы идентификации связей, получаемых из социальных сетей [50].

Цель выпускной квалификационной работы заключается в разработке автоматизированных инструментов, способствующих повышению степени информированности лиц, принимающих решения, о наиболее подверженных социоинженерным атакам цепочках пользователей, за счёт предложения средств выявления наиболее критичных траекторий распространения многоходовых социоинженерных атак злоумышленника и их визуализации на социальном графе сотрудников.

Для достижения цели были поставлены и решены следующие задачи:

- изучение предметной области и релевантных работ по теме исследования;
- разработка методов выявления наиболее вероятных траекторий распространения многоходовых социоинженерных атак;
- разработка метрик оценки критичности траекторий распространения многоходовых социоинженерных атак;
- изучение силы влияния возможных типов взаимоотношений между пользователями, на вероятность распространения социоинженерной атаки;

• разработка алгоритмов, основанных на предложенных методах, и их реализация в прототипе модуля комплекса программ.

Объектом исследования является социальный граф сотрудников компании, построенный на основе данных, извлекаемых из социальных сетей и частично задаваемых экспертно. Такие данные дают одну из возможностей построения оценок вероятности успеха распространения социоинженерной атаки злоумышленника на пользователя.

Предметом исследования являются траектории распространения многоходовых социоинженерных атак, моделируемые на социальном графе пользователей.

Научная новизна. Все результаты, выносимые на защиту, являются новыми. Впервые предложены методы выявления наиболее вероятных траекторий распространения многоходовых социоинженерных атак, методы выявления наиболее критичных траекторий распространения этих атак. Впервые было проведено исследование по изучению силы влияния возможных типов взаимоотношений между пользователями, представленных в социальной сети «ВКонтакте», на вероятность распространения социоинженерной атаки. Впервые были разработаны алгоритмы по выявлению наиболее вероятных траекторий распространения многоходовых социоинженерных атак и выполнена их реализация.

Теоретическая и практическая значимость исследования. Предлагаемые методы, алгоритмы и их реализация позволяют производить в прототипе комплекса программ автоматизированного анализа защищённости пользователей информационных систем от социоинженерных атак поиск наиболее вероятных и наиболее критичных траекторий распространения многоходовых социоинженерных атак. Данная оптимизация способствует дальнейшему развитию исследований по данной тематике, быстрому и эффективному нахождению наиболее уязвимые мест в информационной

системе и, как следствие, позволяет лицам, принимающим решения, производить своевременные меры по повышению уровня защищённости организации.

Методология работы заключается в выявлении наиболее вероятных и наиболее критичных траекторий распространения многоходовых социоинженерных атак, разработке алгоритмов, направленных на нахождение таких траекторий, а также апробации полученных теоретических результатов посредством их реализации в программном комплексе.

Методы включают подходы теории вероятностей, математического анализа, теории графов и объектно-ориентированного программирования. Программная реализация осуществлялась в среде разработки Intellij IDEA 2017 на языке программирования Java.

Положения, выносимые на защиту:

- метод выявления наиболее вероятных траекторий распространения многоходовых социоинженерных атак;
- метрика оценки критичности траекторий распространения многоходовых социоинженерных атак;
- изучение силы влияния возможных типов взаимоотношений между пользователями, на вероятность распространения социоинженерной атаки;
- алгоритмы по выявлению наиболее вероятных траекторий распространения многоходовых социоинженерных атак и их реализация.

Высокая **степень достоверности результатов** выпускной квалификационной работы обеспечивается глубоким и всесторонним анализом исследований по тематике социоинженерных атак, подтверждается согласованностью полученных результатов, а также их успешной апробацией на российских и международных научных конференциях и публикациями в индексируемых изданиях.

Апробация результатов исследования. Результаты исследования в рамках бакалаврской работы были представлены на следующих научных конференциях.

- XVI Санкт-Петербургской международной конференции «Региональная информатика (РИ-2018)» Санкт-Петербург, 2018.
- Нечеткие системы и мягкие вычисления – 2018 (FTI-2018). Ульяновск, 2018.
- «Информационные технологии в управлении» (ИТУ - 2018) Санкт-Петербург, 2018.
- Информационная безопасность регионов России. Санкт-Петербург, 2017.
- ICIT-2019 «Информационно-коммуникационные технологии в науке и производстве».

Выпускная квалификационная работа проводилась в рамках научно-исследовательского проекта, поддержанного грантом РФФИ №18-37-00340 – Методы анализа устойчивости структуры социальных связей пользователей информационной системы к социоинженерным атакующим воздействиям злоумышленника на основе применения генетических алгоритмов.

Публикации. По теме выпускной работы бакалавра было сделано 6 публикаций (2 в изданиях из «Перечня рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание учёной степени кандидата наук, на соискание учёной степени доктора наук», 3 статьи индексируются в Scopus).

Благодарности. Выпускная квалификационная работа бакалавра содержит материалы исследований, выполняемых в рамках государственных заданий СПИИРАН №0073-2018-0001, 0073-2019-0003, а также поддержанных грантами РФФИ: проект №18-01-00626 — Методы представления, синтеза оценок истинности и машинного обучения в алгебраических байесовских сетях и родственных моделях знаний с неопределенностью: логико-вероятностный подход и системы графов; проект

№18-37-00323 — Социоинженерные атаки в корпоративных информационных системах: подходы, методы и алгоритмы выявления наиболее вероятных траекторий, РФФИ №18-37-00340 — Методы анализа устойчивости структуры социальных связей пользователей информационной системы к социоинженерным атакующим воздействиям злоумышленника на основе применения генетических алгоритмов.

Структура и объём работы. Текст работы включает в себя введение, четыре главы, заключение, словарь терминов, список литературы, список иллюстративного материала и приложения. Общий объём выпускной работы бакалавра — 50 страниц.

В главе 1 главе описывается актуальность исследуемой области, также представлено обоснование цели и задач, решаемых в выпускной работе бакалавра.

В главе 2 приводится анализ подходов, решающих схожие задачи и описывается теоретическая часть, послужившая фундаментом для решения поставленных задач, также производится обзор алгоритмов, необходимых для решения поставленных задач.

В главе 3 представлены основные результаты выпускной квалификационной работы, а именно описывается метод для нахождения наиболее вероятных и метрика для оценки критичности траекторий распространения многоходовых социоинженерных атак, приводятся методы квантификации характеристик взаимодействия пользователей.

В главе 4 описаны процесс разработки и структура программных модулей: «Анализатора критичных траекторий» и «Построения графа».

1. Теоретические основы

В данной главе описывается актуальность исследуемой области, также представлено обоснование цели и задач, решаемых в выпускной квалификационной работе бакалавра.

1.1. Актуальность задач защиты пользователей от социоинженерных атак

В настоящее время проблема анализа защищённости пользователей информационных систем от социоинженерных атак является одной из актуальных, что подтверждается количеством инцидентов и интенсификацией роста убытков от них [26, 27, 32, 41, 43, 47]. Одним из подтверждений актуальности исследуемой тематики может служить информация, предоставляемая крупными российскими компаниями, свидетельствующая о применении методов социальной инженерии в 80% случаев от числа всех инцидентов нарушения безопасности [48]. Данная информация также отражена и в отчётах компаний, специализирующихся на разработке систем защиты от кибератак [40, 44]. Актуальность данного направления находит подтверждение и у экспертов в области информационной безопасности [32]. Согласно их прогнозам [54] в ближайшие несколько лет ожидается рост активности злоумышленников-социоинженеров.

Помимо сотрудников организаций жертвами социоинженерных атак также становятся и частные лица. Последние годы в России в этой роли оказывались 1,2 млн человек ежегодно [45]. В первом полугодии 2018 года у жертв в среднем было похищено 5000 рублей, что составляет 12% среднемесячной зарплаты россиянина [45]. По классификации, представленной в монографии [39], особую группу социоинженерных атак составляют фишинг-атаки, кроме того, отмечается их существенный рост за последнее время. Данный факт также находит

подтверждение в отчёте компании Group-IB [53]: при помощи веб-фишинга хакерам удалось украсть около $4,2 млн. При этом в среднем за день совершается более 1,2 тысяч социоинженерных атак [53].

В связи с этим актуальной видится задача защиты пользователей информационных систем от социоинженерных атак. Одной из мер, способствующих обеспечению защиты пользователей информационных систем от социоинженерных атак, является анализ защищенности. Общая цель направления исследований заключается в повышении уровня информационной безопасности организации за счет разработки автоматизированных средств анализа защищённости пользователей информационных систем от социоинженерных атак.

1.2. Многоходовые социоинженерные атаки

Для решения этой проблемы необходимо, в частности, строить оценки защищённости пользователей от прямых и многоходовых социоинженерных атак злоумышленника. Многоходовая социоинженерная атака отличается от прямой (одноходовой) тем, что совершается через цепочку пользователей, а не происходит опосредованно. Многоходовые социоинженерные атаки, как правило, могут распространяться по нескольким траекториям. Оценки вероятности успеха распространения многоходовой атаки злоумышленника на пользователя по разным траекториям обычно имеют отличающиеся значения. В связи с этим, необходимо выявить траектории, успешное прохождение по которым произойдёт с большей вероятностью. Для этого предлагается производит анализ на социальном графе взаимодействия сотрудников.

Подход к оценке защищённости пользователей при многоходовых социоинженерных атаках, а также к расчёту оценок вероятности распространения атаки от пользователя к пользователю представлен в [2, 39, 52]. Отметим, что в [2] рассмотрены неориентированные графы, однако оценка вероятности распространения атаки от пользователя к пользователю в прямом и обратном

направлении может быть разной. Также в [2] не рассматриваются вопросы поиска наиболее критичных траекторий распространения атаки. Как правило, существует несколько возможных траекторий развития атаки от одного пользователя к другому и вероятности успеха распространения атаки по каждой из них будут иметь разные значения. В связи с этим, актуальной видится проблема выявления наиболее вероятных и наиболее критичных траекторий распространения.

Данная работа является продолжением указанного выше общего исследования и опирается на полученные ранее результаты [2, 30, 38, 39, 51, 52]. А именно анализ траекторий осуществляется в уже построенном и размеченном графе [39], отражающем социальные связи между пользователями системы и вероятности успеха перехода злоумышленника от пользователя к пользователю. Предполагается, что данное решение будет способствовать повышению оперативности выявления наиболее незащищенных звеньев информационной системы, в следствие чего могут быть приняты своевременные меры по их защите.

1.3. Цели и задачи

Приводимые выше сведения обосновывают актуальность проблематики социоинженерных атак. Вместе с тем одним из малоизученных мест по данной тематике видится исследование сложных социоинженерных атак, осуществляемых через цепочку пользователей. Такие атаки называются многоходовыми социоинженерными атаками [2]. Атаковать целевого пользователя при этом можно через разные цепочки, и оценки вероятности прохождения по ним будут отличаться. Если представить сотрудников организации в виде социального графа, то можно говорить о разных траекториях реализации многоходовых социоинженерных атак. При этом оценки вероятности успешного прохождения этих траекторий отличаются. В связи с чем, возникает необходимость выявления наиболее вероятной траектории распространения многоходовой социоинженерной атаки или совокупности таких траекторий.

Кроме того, важно учитывать, что от реализации разных траекторий организация несет отличающиеся по размеру убытки. Таким образом, существенно выявлять не только наиболее вероятные траектории, но разработать подход к идентификации наиболее критичных траекторий. В связи с этим была выдвинута следующая цель, достигаемая в выпускной квалификационной работе бакалавра:

Цель выпускной квалификационной работы заключается в разработке автоматизированных инструментов, способствующих повышению степени информированности лиц, принимающих решения, о наиболее подверженных социоинженерным атакам цепочках пользователей, за счёт предложения средств выявления наиболее критичных траекторий распространения многоходовых социоинженерных атак злоумышленника и их визуализации на социальном графе сотрудников.

Для достижения цели были поставлены и решены следующие задачи:

- разработка методов выявления наиболее вероятных траекторий распространения многоходовых социоинженерных атак;
- разработка метрик оценивания критичности траекторий распространения многоходовых социоинженерных атак;
- изучение силы влияния возможных типов взаимоотношений между пользователями, на вероятность распространения социоинженерной атаки;
- разработка алгоритмов, основанных на предложенных методах, и их реализация в прототипе модуля комплекса программ.

Выводы по главе. В данной главе была показана актуальность исследуемой области, описаны многоходовые социоинженерные атаки. Представлено обоснование цели и задач, решаемых в выпускной работе бакалавра.

2. Используемые подходы и методы

В данной главе приводится анализ подходов, решающих схожие задачи и описывается теоретическая часть, послужившая фундаментом для решения

поставленных задач, приводится обзор существующих алгоритмов по поиску кратчайшего пути в графе.

2.1. Релевантные работы

Заделом для данного исследования послужили работы [2, 5] в которых описаны подходы к оценке защищенности пользователей информационных систем от прямых и многоходовых социоинженерных атак. Исследования по повышению уровня защищённости пользователя были представлены в [18]. В данной работе авторы представили многоуровневую модель оценки уязвимости пользователей к социоинженерным атакам, базирующуюся на трёх основных элементах: способ связи, состояние системы и сценарий атаки, также проведены эксперименты по анализу эффективности разработанной системы. Подход, описанный в [36], основывается на преобразовании требований безопасности в элементы обучающей игры, в результате использования разработанной игровой программы пользователи информационной системы могут распознать основные сценарии социоинженерных атак.

Часть рассмотренных исследований была направлена на изучение поведения пользователей в социальных сетях. Результатом исследования [4, 14, 19, 25] являются эмпирические оценки восприимчивости пользователей к социоинженерным атакам. Сделанные на основе упомянутых исследований выводы могут быть полезны для оптимизации процесса анализа контента, извлекаемого из социальных сетей. Схожую направленность имеет исследование [23], в нём приводятся факторы, влияющие на соблюдение сотрудниками политики безопасности. В [3] исследуются факторы, влияющие на уязвимости пользователей и причины подверженности социоинженерным атакам.

Исследование, основывающиеся на анализе текста и направленное на выявление фишинговых писем, представлено в [6]. В исследовании [7] изучается вопрос защиты от прогнозирования скрытых конфиденциальных данных

пользователей социальных сетей, также исследуется зависимость между открытой и скрытой информацией в социальном профиле пользователя, результатом исследования является разработка способа защиты. Схожая проблема поднимается в источнике [16], подход предлагаемый авторами данной статьи базируется на автоматизированном сборе информации из открытых источников, её анализе и выявлении критических с точки зрения безопасности мест. Результатом [21] стала разработка web-сервиса, основанного на интеллектуальном анализе данных и позволяющего распознавать угрозы безопасности в социальной сети Twitter. Авторы [8, 20, 37] производят анализ неявного социального графа (графа, формируемого на основе данных о «друзьях» в социальной сети) с целью обнаружения аномального поведения, а также подозрительных и ложных учётных записей. В [1] поднимается проблема нарушения конфиденциальности данных социальных сетей и предлагается ряд методов для её решения. Исследование [11] предоставляет комплексный обзор ключевых исследований в области конфиденциальности информации за последние 40 лет.

2.2. Алгоритмы поиска кратчайшего пути на социальном графе сотрудников компании

В ходе проведения исследований было выявлено, что для решения поставленных задач необходимо уметь осуществлять поиск кратчайшего пути на социальном графе сотрудников компании. Пусть n — число вершин в социальном графе (число сотрудников), m — число дуг в социальном графе. Для решения задачи поиска кратчайшего пути на социальном графе были рассмотрены алгоритмы Беллмана–Форда, Левита, Флойда–Уоршелла, Дейкстры и его модификации, топологическая сортировка, A* [12, 22, 29].

Алгоритмы Левита и Флойда–Уоршелла имеют высокую вычислительную сложность в контексте социального графа пользователей, поэтому их применение

нецелесообразно. Для применения алгоритма топологической сортировки исходный граф должен быть ацикличным. Социальный граф сотрудников компании в большинстве случаев не обладает этим свойством, в связи с чем указанный алгоритм не может быть применен. Алгоритм A* удобен тем, что осуществляет поиск кратчайшего расстояния только между двумя вершинами, а не между всеми. Однако трудность использования данного алгоритма заключается в подборе правильной эвристической функции. Кроме того, алгоритм A* использует большой объём памяти при работе, в связи с чем его использование нецелесообразно. Алгоритм Дейкстры подходит для задачи поиска кратчайшего пути на социальном графе и обладает вычислительной сложностью O(n2). При этом оптимальная сложность для алгоритмов, основанных на алгоритме Дейкстры, составляет O(nlogn+m) и достигается при представлении данных в виде куч Фибоначчи. Однако константы, скрытые в асимптотических оценках трудоемкости упомянутой модификации, зачастую на практике велики. С другой стороны, данные можно хранить в двоичной куче, тогда сложность составит O(nlogn+mlogn). Но заметим, что время работы модификаций сократится по сравнению с классическим алгоритмом Дейкстры только при условии $m \ll n2$, то есть в случае разреженного графа. В рамках нашей задачи не всегда предполагается работа с разреженными социальными графами.

Также в случае, если число дуг в графе меньше числа вершин ($m<n$), то для поиска наикратчайшего пути используется алгоритм Беллмана–Форда. Его вычислительная сложность O(mn), и в этом случае она меньше, чем у алгоритма Дейкстры.

2.3. Оценка вероятности перехода социоинженерной атаки между двумя пользователями

Анализ возможных траекторий предлагается производить на ориентированном социальном графе сотрудников компании. Под социальным графом сотрудников

компании будем понимать граф, вершины которого соответствуют сотрудникам компании, а рёбра — связям между сотрудниками. Формализуем описанное. Пусть дан граф:

$$G = (U, E), \tag{1}$$

где $U = \{User_i\}_{i=1}^{n}$ — множество вершин (пользователей), $E = \{(u_i, u_j, p_{i,j})\}_{1 \le i,j \le n, i \ne j}$ — множество упорядоченных троек с заданной оценкой вероятности распространения атаки от пользователя к пользователю — $p_{i,j}$. Заметим, что равенство $p_{i,j}$ и $p_{j,i}$ не предполагается. То есть вероятность распространения атаки от первого пользователя ко второму может отличаться от вероятности распространения атаки в другую сторону — от второго к первому.

Отметим, что оценка вероятности распространения социоинженерной атаки от пользователя к пользователю, согласно [38] рассчитывается следующим образом:

$$p_{i,j} = 1 - \prod_t \left(1 - p_t^{i,j}\right)^{n_t}, \tag{2}$$

где $p_t^{i,j}$ —оценка вероятности успеха социоинженерной атаки злоумышленника на пользователя по t-ой связи, n_t — число эпизодов. В рассматриваемом частном случае модели, учитывающей сведения, извлекаемые из социальных сетей, $p_{i,j} > 0$, рёбра, где оценка вероятности $p_{i,j} = 0$ исключаются из итогового социального графа.

Задача поиска наиболее вероятной траектории многоходовой социоинженерной атаки от $User_i$ до $User_j$ сводится к задаче нахождения в графе элементарного пути (простого и без циклов) между этими вершинами. Причём путь должен быть таким, что произведение оценок вероятностей переходов от

пользователя к пользователю, входящих в него, максимально:

$$p_{ml} = \underset{Trajectories}{\mathrm{Max}}\left(p_m \prod_{i,j} p_{ij} \right), \; p_{ml} = \underset{Trajectories}{\mathrm{Argmax}}\left(p_m \prod_{i,j} p_{ij} \right).$$

Будем называть оценку вероятности успеха многоходовой социоинженерной атаки, которая представляет собой произведение оценок вероятностей распространения атаки от пользователя к пользователю и прямой атаки на первого пользователя, длиной пути.

Выводы по главе. Приведён анализ подходов, решающих схожие задачи и описывается теоретическая часть, послужившая фундаментом для решения поставленных задач. Также приведён обзор существующих алгоритмов по поиску кратчайшего пути в графе.

3. Траектории реализации многоходовых социоинженерных атак

В данной главе описываются метрики для нахождения наиболее вероятных и наиболее критичных траекторий распространения многоходовых социоинженерных атак, а также представлены методы квантификации характеристик взаимодействия пользователей.

3.1. Квантификация интенсивности взаимодействия пользователей

В ходе процесса исследования возникла задача квантификации одной из характеристик интенсивности взаимодействия пользователей, являющихся узлами в социальном графе сотрудников. А именно, требовалось сопоставить рёбрам социального графа их веса. Произвести это предлагается при помощи численной оценки взаимосвязей пользователей, указанных в социальной сети «ВКонтакте» (например, родственники, лучшие друзья, брат/сестра и т.д.). Для этого были рассмотрены следующие типы классификации друзей, предлагаемые социальной

сетью «ВКонтакте»: лучшие друзья, родственники, коллеги, друзья по вузу, друзья по школе, с ним/ней встречаюсь, жених/невеста, муж/жена, в гражданском браке, влюблён/а в, все сложно, дедушка или бабушка, родитель, брат или сестра, сын или дочь, внук или внучка. Данные категории были разбиты на три группы:

1. Публичный список друзей: лучшие друзья, родственники, коллеги, друзья по вузу, друзья по школе.
2. Основная информация в аккаунте: с ним/ней встречаюсь, жених/невеста, муж/жена, в гражданском браке, влюблён/а в, все сложно.
3. Семейное положение: дедушка или бабушка, родитель, брат или сестра, сын или дочь, внук или внучка.

Для удобства внесения ответов и обработки результатов была разработана веб-страница с опросом (в категории друзья).

В качестве вариантов ответа респондентам предлагалось установить бегунок на шкале от 0 до 100 с шагом 1, где значение 100 соответствует максимальной вероятности выполнения некоторого действия, если об этом попросить пользователь, относящийся к данной категории. 0 соответствует тому, что просьба, поступившая от пользователя указанной категории не будет выполнена ни при каких условиях. Вопрос звучал следующим образом: «Представьте следующую ситуацию: Вам пришло приглашение вступить в группу ВКонтакте. Оцените, пожалуйста, с какой вероятностью Вы бы откликнулись на эту просьбу, если бы Вам пришло приглашение от человека, который отмечен у Вас во ВКонтакте как:».

В опросе приняли участие 145 человек, среди которых было 88 девушек и 57 молодых людей. Их средний возраст составил 22 года, медиана по возрасту – 20 лет. Большая часть респондентов являются студентами ведущих вузов России по техническим, гуманитарным и управленческим специальностям.

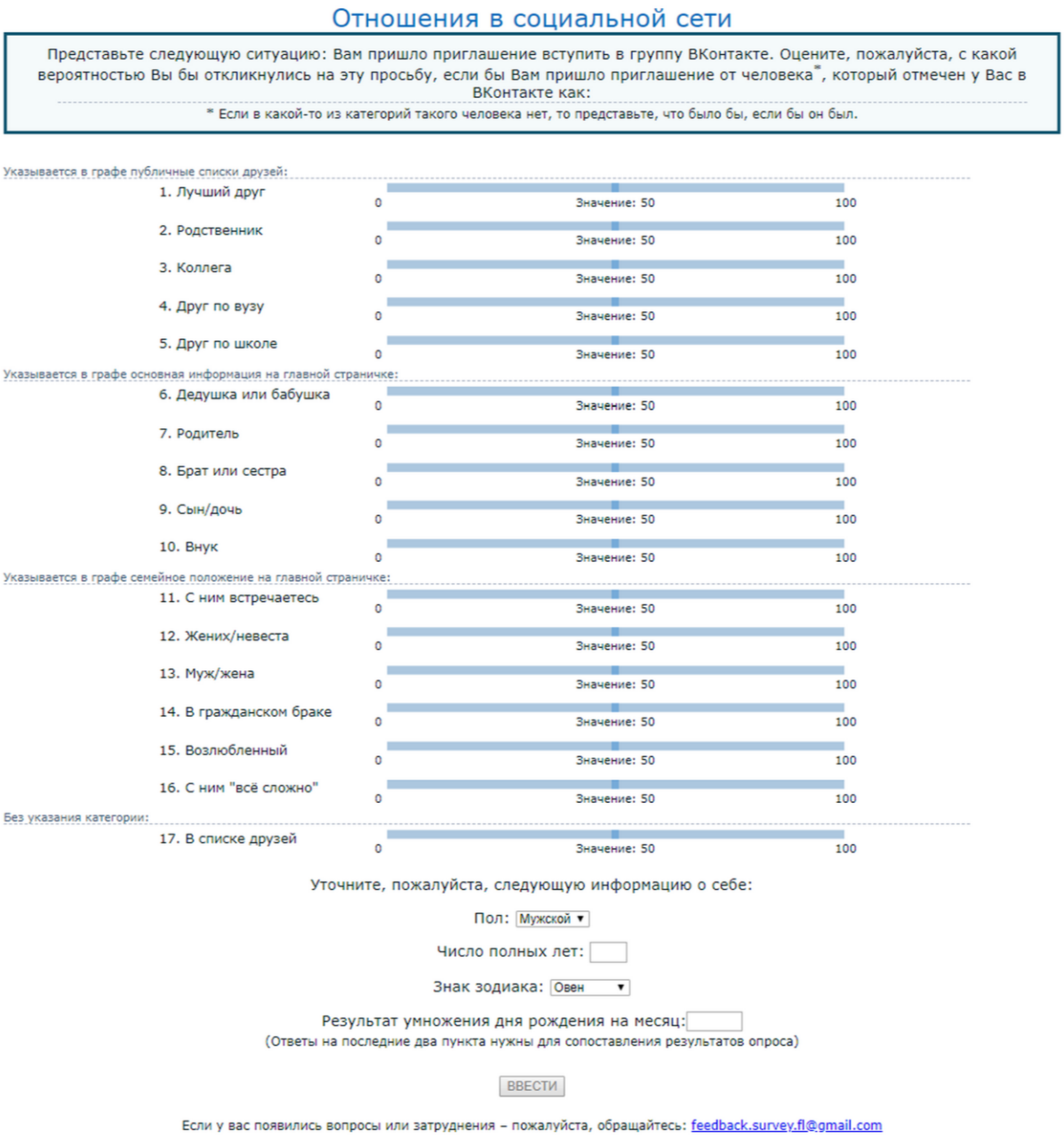


Рисунок 1 — Скриншот web-страницы с опросом

В ходе проведения первичного анализа результатов опроса было отмечено, что ряд пользователей не различали две или более категории. То есть часто встречались ответы, в которых нескольким или даже всем категориям в группе были присвоены одинаковые оценки степени готовности отреагировать на просьбу. Например, среди родственников, отмеченных пользователем в основной информации аккаунта, существенная часть респондентов отметили одинаковые

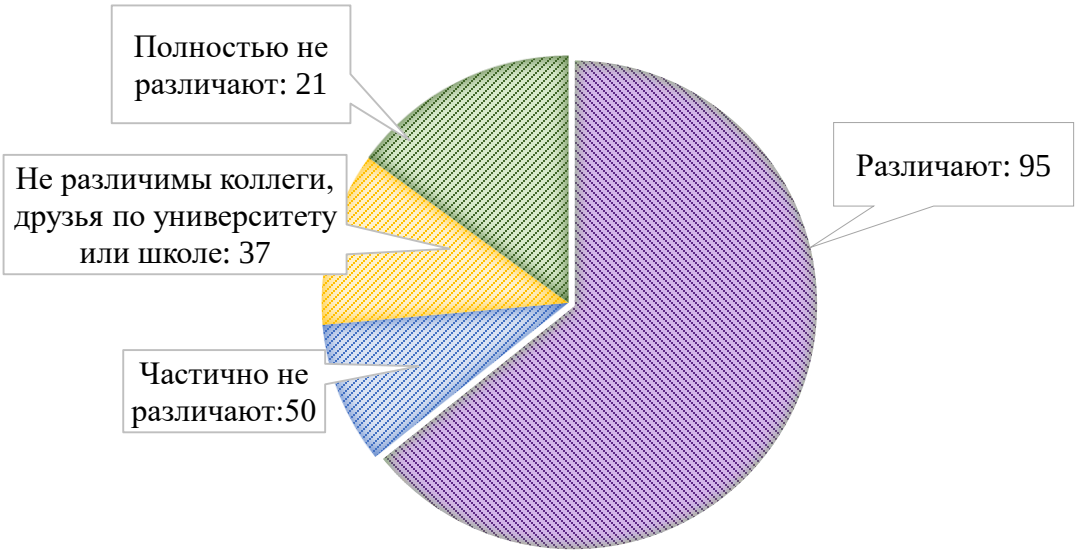
оценки готовности откликнуться на просьбу. Получилось, что позитивно отреагируют на просьбу бабушки/дедушки, родителя, брата/сестры с одинаковой вероятностью. В связи с этим, было решено посчитать число респондентов, расположивших пользователей из разных категорий в одинаковом порядке.

Для автоматизации данного процесса была разработана программа на языке C# с использованием библиотеки Microsoft Excel Object Library. В качестве входных данных программы служит документ Microsoft Excel с результатами опроса. На выходе подаётся также документ Microsoft Excel с добавленными страницами, по одной на каждую подгруппу. Каждая из таких страниц содержит сгенерированный порядок ответов, расшифровку данного порядка и число раз, которое такой порядок встретился в упорядоченных ответах респондентов по каждой подгруппе. Пример результата работы программы по одной из групп представлен в таблице 1.

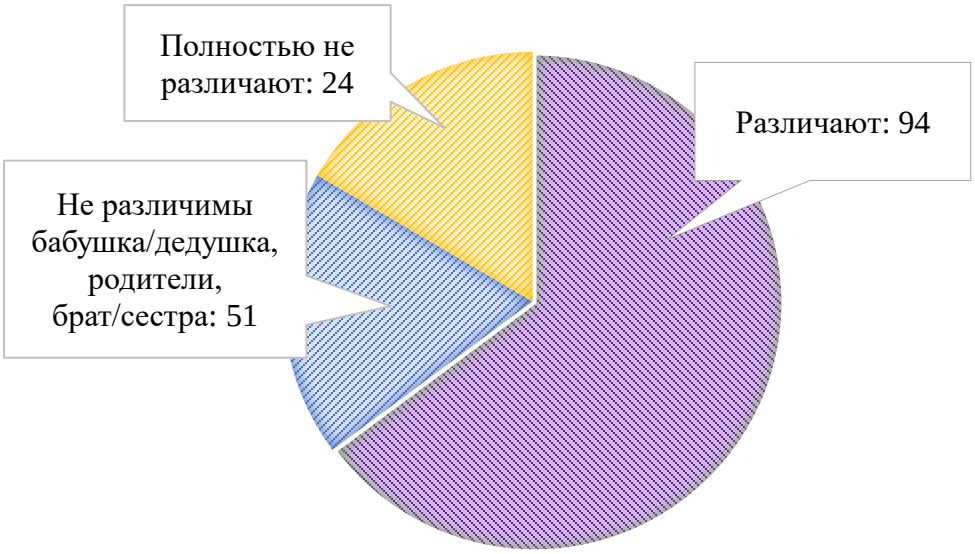
Таблица 1 — Пример части обработанных результатов опроса

Порядок	Частот-ность	Расшифровка порядка					
[123456]	24	Семья	Дедушка/ бабушка	Родители	Брат/ сестра	Ребёнок	Внук/ внучка
[23456]1	7	Дедушка/ бабушка	Родители	Брат/ сестра	Ребёнок	Внук/ внучка	Семья
1[23456]	5	Семья	Дедушка/ бабушка	Родители	Брат/ сестра	Ребёнок	Внук/ внучка
[1234][56]	3	Семья	Дедушка/ бабушка	Родители	Брат/ сестра	Ребёнок	Внук/ внучка
[2356]41	2	Дедушка/ бабушка	Родители	Ребёнок	Внук/ внучка	Брат/ сестра	Семья
[61234]5	2	Внук/ внучка	Семья	Дедушка/ бабушка	Родители	Брат/ сестра	Ребёнок
1[2364]5	2	Семья	Дедушка/ бабушка	Родители	Внук/ внучка	Брат/ сестра	Ребёнок
[12356]4	2	Семья	Дедушка/ бабушка	Родители	Ребёнок	Внук/ внучка	Брат/ сестра
1[6234]5	2	Семья	Внук/ внучка	Дедушка/ бабушка	Родители	Брат/ сестра	Ребёнок
[41256]3	2	Брат/ сестра	Семья	Дедушка/ бабушка	Ребёнок	Внук/ внучка	Родители
231654	1	Дедушка/ бабушка	Родители	Семья	Внук/ внучка	Ребёнок	Брат/ сестра

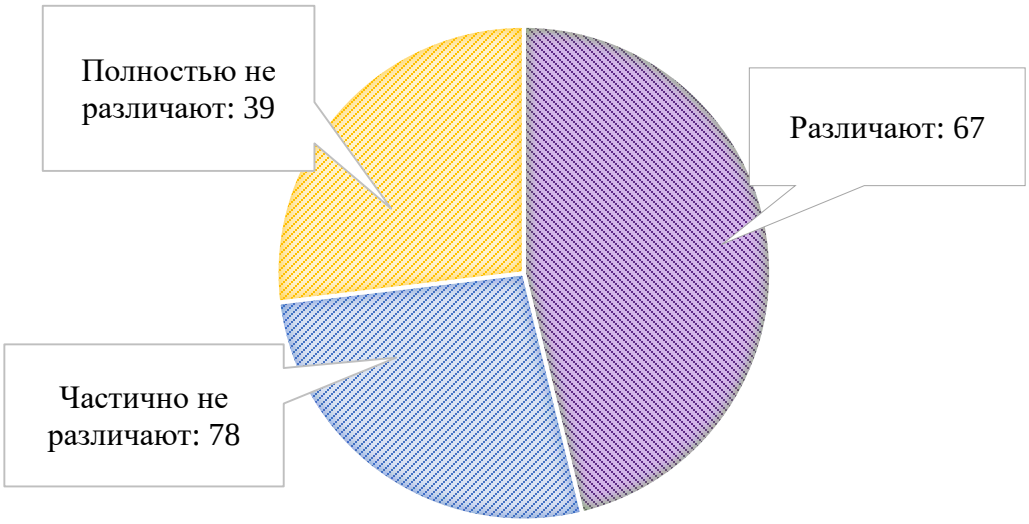
В результате исследования полученных соотношений было выявлено, что во всех категориях более, чем для трети респондентов указанные взаимосвязи практически неразличимы. В категории друзья 50 человек практически не разделяют связи, из них 37 человек не разделяют связи типа: коллеги, друзья по университету, друзья по школе и 21 — полностью не различают связи данной категории (Рис. 2(а)). В категории родственники 51 респондент не разделяют связи типа: дедушка/бабушка, родители и брат/сестра, из них 24 человека присвоили одинаковые оценки всем связям данной категории (Рис. 2(б)). Безразличных по дифференциации в категории семейное положение более половины опрошенных — 78 человек, из них 39 человек полностью не дифференцируют связи данной категории (Рис. 2(в)).



(а) в категории друзья



(б) в категории родственники



(в) в категории семейное положение

Рисунок 2 — Соотношение результатов опроса

Основная гипотеза, выдвинутая по результатам данного исследования и требующая проверки, заключается в том, что оценки степени готовности откликнуться на просьбу вступить в сообщество для разных групп взаимоотношений различны, но мало отличаются внутри группы.

Полученные результаты, демонстрирующие отсутствие дифференциации значений внутри групп типов взаимоотношений, являются существенными, но в то же время требуется более глубокое изучение порядков, которые можно отследить в ответах ряда респондентов. Вместе с этим для массовых исследований нельзя игнорировать то, что были рассмотрены далеко не все возможные сочетания, а реализация расстановки отношений может быть более многообразной. Ещё одним направлением дальнейших исследований может послужить отслеживание динамики по возрастным категориям. Планируется проведение повторного опроса тем же инструментом с целью проверки устойчивости оценивания связей. Для интерпретации результатов целесообразным видится использование метода Н.В. Хованова [46, 55], который позволяет исследовать отношения порядков друг к другу. Предполагается, что выбор конкретной численной позиции случаен, а разработанный инструмент помогает респондентам отобразить ранжирование связей. Вероятно, результаты, полученные с применением метода Хованова, окажутся более устойчивым, чем численные данные.

Указанные результаты в перспективе позволят явно задавать параметры моделей для построения оценок распространения многоходовых социоинженерных атак.

3.2. Выявление наиболее вероятной траектории распространения социоинженерной атаки между двумя пользователями

Для упрощения, не умаляя общности, рассмотрим граф $G = (U, E')$, где $U = \{\text{User}_i\}_{i=1}^{n}$ — множество вершин, $E' = \left\{\left(u_i, u_j, \frac{1}{p_{i,j}}\right)\right\}_{1 \le i,j \le n, i \ne j}$ — множество упорядоченных троек, где каждой паре пользователей u_i, u_j сопоставлено

число $\frac{1}{p_{i,j}}$. Согласно [38] вероятность успеха прохождения многоходовой социоинженерной атаки от пользователя m до пользователя l может быть рассчитана как: $p_{ml} = p_m \prod_{i=m}^{l-1} p_{i,i+1}$. Заметим, что в этом случае если $p_{i,j} \geq p_{l,k}$, то $\frac{1}{p_{i,j}} \leq \frac{1}{p_{l,k}}$, а длина пути будет вычисляться следующим образом $\frac{1}{p_{ml}} = \frac{1}{p_m} \prod_{i=m}^{l-1} \frac{1}{p_{i,i+1}}$, где p_{ml} — оценка вероятности успеха прохождения атаки от пользователя m до пользователя l, p_m — оценка вероятности успеха прямой социоинженерной атаки злоумышленника на пользователя, $p_{i,i+1}$ — соответствующая оценка вероятности распространения атаки на пользователя через другого пользователя. Таким образом, от задачи поиска пути с максимальной длиной перейдём к задаче поиска пути с минимальной длиной.

Чтобы применить алгоритмы нахождения минимального пути, необходимо произвести ряд преобразований. Согласно основному логарифмическому тождеству $\frac{1}{p_{ij}} = e^{\log \frac{1}{p_{ij}}}$. Тогда длина пути будет рассчитываться следующим образом $\frac{1}{p_{ml}} = \frac{1}{p_m} \prod_{i=m}^{l-1} \frac{1}{p_{i,i+1}} = e^{\log \frac{1}{p_m}} \prod_{i=m}^{l-1} e^{\log \frac{1}{p_{i,i+1}}} = \exp\left\{ \log \frac{1}{p_m} + \sum_{i=m}^{l-1} \log \frac{1}{p_{i,i+1}} \right\}$.

Поскольку оценка успеха прямого социоинженерного атакующего воздействия на пользователя m — p_m — будет одинакова для всех траекторий, начинающихся с пользователя m, то задача сводится к поиску пути, в котором $-\sum_{i=m}^{l-1} \log p_{i,i+1}$ — минимальна среди всех возможных траекторий, начинающихся с

пользователя m и заканчивающихся пользователем l, или, что то же $\sum_{i=m}^{l-1} \log p_{i,i+1}$ максимальна. Таким образом, задача представляет собой стандартный поиск наикратчайшего пути в ориентированном графе без рёбер отрицательного веса.

3.3. Обобщение задачи

В связи с тем, что выявление наиболее вероятных траекторий без оценки ущерба от их реализации не даёт необходимой информации для принятия мер по повышению уровня безопасности, в ходе дальнейших исследований возник вопрос о поиске наиболее критичных траекторий атак не с точки зрения вероятности поражения пользователя или документа, а с точки зрения ожидаемого ущерба. Такая характеристика может быть построена на основе анализа возможностей и прав доступа пользователя к документам различной критичности. Отметим, что критичные документы в информационной системе могут иметь разные уровни критичности и, соответственно, их компрометация будет приводить к различающимся по размеру ущербу. Обычно критичные документы разделяют на группы по уровню критичности. К первой группе относятся наиболее критичные документы, к последней наименее критичные. Рассмотрим подходы к распределению прав доступа в информационных системах к критичным документам разных уровней.

Первый подход к распределению прав доступа к критичным документам заключается в том, что критичные документы разбиты по группам, исходя из их уровня критичности, и каждый пользователь информационной системы имеет доступ к документам какого-то одного уровня критичности (Рисунок 3). В этом случае модели критичных документов и пользователя информационной системы могут быть представлены следующим образом (Таблица 2).

Но чаще пользователи имеют доступ не только к документам какого-то одного уровня критичности, но и к документам уровня критичности ниже

(Рисунок 4). Т.е., когда критичные документы разбиты по группам по уровню критичности, и пользователи имеют доступ к критичным документам своего уровня критичности и документам всех уровней ниже. Наиболее распространённой является модель распределения по уровням доступа, когда пользователи имеют доступ не ко всем документам определённого уровня критичности, а только к части из них (Рисунок 5). Т.е., когда документы разбиты по уровням критичности, а пользователи имеют доступ к определённому количеству критичных документов каждого уровня критичности.

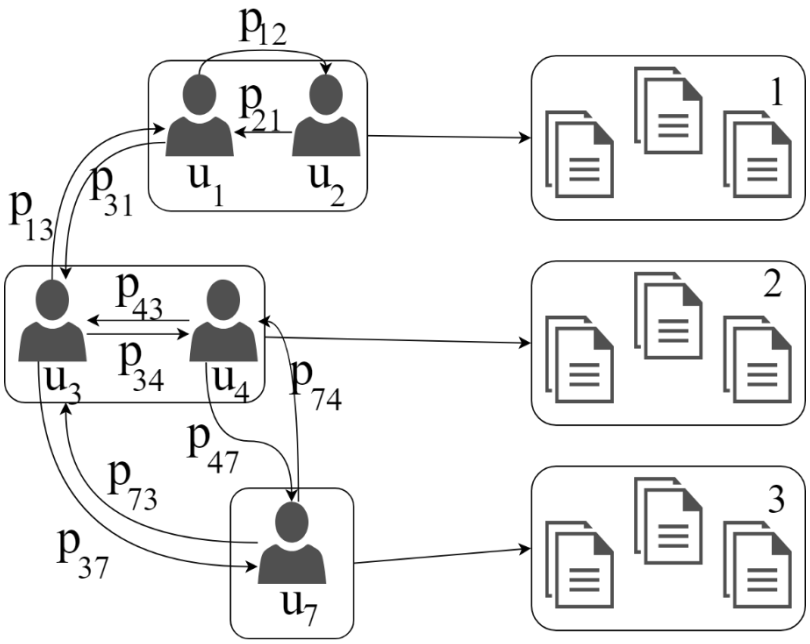


Рисунок 3 — Распределение прав доступа пользователей, при котором каждый пользователь имеет доступ к документам какого-то одного уровня критичности.

Таблица 2 — Модели пользователя и критичных документов

№	Название	Представление	Комментарий
1	Критичный документ	cd(id,lc)	Критичные документ характеризуется идентификационным номером и уровнем критичности.
2	Пользователь	users(id,lc)	Пользователь в системе имеет идентификационный номер, через который связан с другими атрибутами, и уровень доступа к критичным документам.

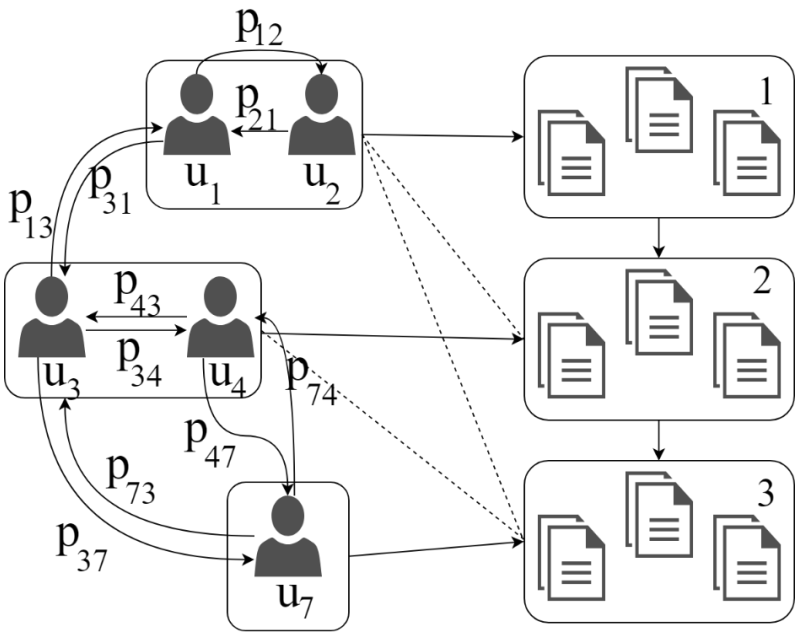


Рисунок 4 — Распределение прав доступа пользователей: каждый пользователь имеет доступ к документам определённого уровня критичности и ко всем документам нижнего уровня.

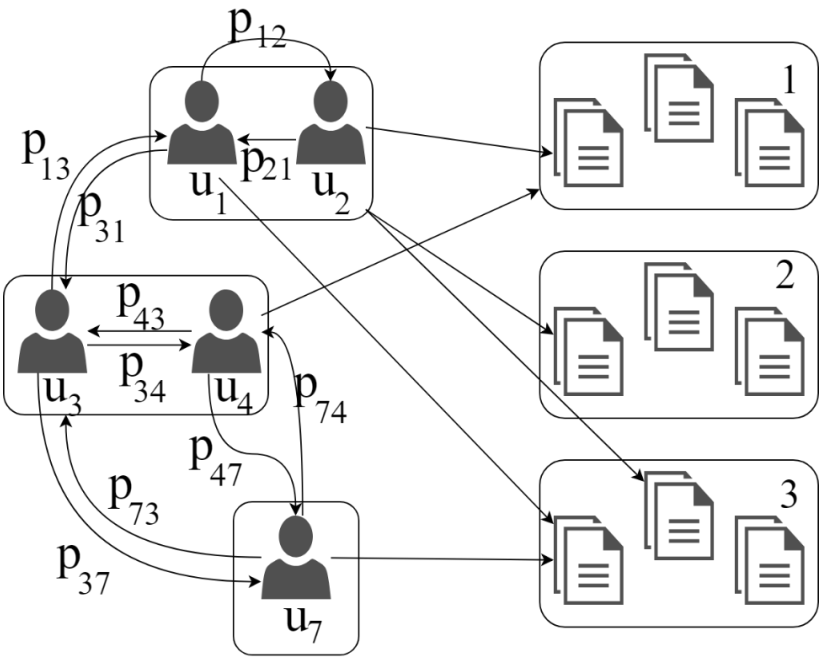


Рисунок 5 — Распределение прав доступа пользователей: каждый пользователь имеет доступ к определённым документам разных уровней критичности.

В работе рассматривается задача выявления наиболее критичной траектории распространения социоинженерной атаки для информационной системы, в которой права доступа распределены так, что критичные документы разбиты по

группам, исходя из их уровня критичности, и каждый пользователь информационной системы имеет доступ к документам какого-то одного уровня критичности.

3.4. Подход к идентификации наиболее критичной траектории

Как было выявлено ранее [39] вероятность прохождения атаки от пользователя m до пользователя l — это

$$p_{ml} = \underset{Trajectories}{\text{Max}} \left(p_m \prod_{i,j} p_{ij} \right),$$

где $Trajectories = \{(User_m, E_{i_1}, ..., E_{i_k}, User_l)\}_{i_1,...,i_k}$ — множество всевозможных траекторий распространения многоходовой социоинженерной атаки между заданными пользователями, p_m — оценка вероятности успеха прямой социоинженерной атаки злоумышленника на пользователя m, p_{ij} — соответствующая оценка вероятности распространения атаки на пользователя j через пользователя i.

Как было отмечено выше, выявление наиболее вероятных траекторий без оценки ущерба от их реализации не даёт нам необходимой информации, которая позволила бы принимать превентивные таргетированные меры, способствующие повышению уровня информационной безопасности в организации. В связи с этим, необходимо перейти от выявления наиболее вероятных траекторий к выявлению наиболее критичных траекторий. Наиболее критичной траекторией будем называть наиболее вероятную траекторию реализации социоинженерной атаки, которая принесёт максимальный ущерб организации. Для оценки критичности траекторий предлагается ввести соответствующую метрику, которая может быть формализована следующим образом

$$\mathrm{ct}_{ml} = p_{ml} \cdot \mathrm{loss}(l, \mathrm{lc}), \quad (1)$$

где ct_{ml} — оценка критичности траектории между пользователями m и l, p_{ml} — максимальная оценка вероятности прохождения социоинженерной атаки между данными пользователями, а $\mathrm{loss}(l, \mathrm{lc})$ — потенциальный ущерб организации при компрометации критичных документов, доступных пользователю l, уровня критичности lc. Таким образом, необходимо найти траекторию $\mathrm{ct}: \mathrm{ct} = \underset{\mathrm{User}_{m,l} \in U}{\mathrm{Max}} (\mathrm{ct}_{ml})$.

Простейшим вариантом нахождения такой траектории является расчёт и ранжирование всевозможных вариантов значений ct_{ml} для разных m и l. Однако указанный подход является ресурсозатратным. Для снижения ресурсозатратности можно двигаться в сторону сужения области перебора значений оценок вероятностей. Подобным фильтром может выступать задание нижнего порога для оценок вероятностей прохождения траекторий. А также задание порогового уровня критичности документа, убытка при его компрометации, при которых итоговое значение критичности траектории будет минимальным.

Для большей наглядности предложенного подхода рассмотрим пример. Пусть дан социальный граф, состоящий из трех сотрудников компании, с указанными оценками вероятностей распространения социоинженерной атаки от пользователя к пользователю. Также представим, что в информационной системе содержатся три критичных документа трех разных уровней критичности. К каждому из таких документов имеет доступ один пользователь. Зададим следующие значения для $\mathrm{loss}(l, \mathrm{lc})$: $\mathrm{loss}(1,1) = 1$, $\mathrm{loss}(2,2) = 2$, $\mathrm{loss}(3,3) = 3$. Т.е. документ 1 имеет уровень критичности 1, документ 2 имеет уровень критичности 2, документ 3 — 3. Уровень 1 соответствует наименьшему уровню критичности,

уровень 3 — наивысшему. Рассчитаем оценки критичности траекторий между пользователями для указанных условий согласно (1):

$$c_{12} = 0.9 \cdot 0.8 \cdot 2 = 1.44,$$
$$c_{13} = 0.9 \cdot (0.8 \cdot 0.6) \cdot 3 = 1.296,$$
$$c_{21} = 0.67 \cdot 0.9 \cdot 1 = 0.603,$$
$$c_{23} = 0.9 \cdot 0.6 \cdot 3 = 1.62,$$
$$c_{31} = 0.23 \cdot (0.8 \cdot 0.9) \cdot 1 = 0.1656,$$
$$c_{32} = 0.23 \cdot 0.8 \cdot 2 = 0.368.$$

Таким образом, $ct = 1.62,$ что соответствует атаке на критичный документ уровня критичности 3 через пользователя 3, атакованного посредством пользователя 2. Отметим, что для примера были рассмотрены простейшие значения функции $\mathrm{loss}(l, \mathrm{lc})$, также был использован простейший пример информационной системы с тремя пользователями, критичными документами и уровнями критичности. Возможно, дальнейшие исследования покажут необходимость изменения принципов задания и распределения уровней критичности документов, данные величины могут быть заданы экспертами и выражены более сложным образом.

Выводы по главе. Были введены метод нахождения наиболее вероятных и метрики оценки критичности траекторий распространения многоходовых социоинженерных атак, а также представлены методы квантификации характеристик взаимодействия пользователей.

4. Программная реализация

В данной главе описывается структура программных модулей: «Анализатора критичных траекторий» и «Построения графа».

4.1. Структура программного модуля

Разработанный программный модуль является частью программного комплекса [38, 39], представленного на Рисунок 6. Программный комплекс реализует комплексный анализ информационной системы компании на предмет выявления наиболее уязвимых мест к воздействию злоумышленников-социоинженеров. Работа комплекса включает в себя агрегацию данных из социальных сетей, анализ данных о психологических особенностях сотрудников компании, обработку полученной информации, построение на её основе социального графа и его дальнейшего анализа.

Разработка программных модулей велась на языке программирования Java в программной среде Intellij Idea 2017. В качестве исходной структуры для исследования была взята структура *Social Graph* [39], включающая в себя следующие два объекта *users* и *connections*. Для связи с разработанными раннее модулями и для предоставления удобного доступа в дальнейшем к разрабатываемым модулям была использована технология Apache Maven.

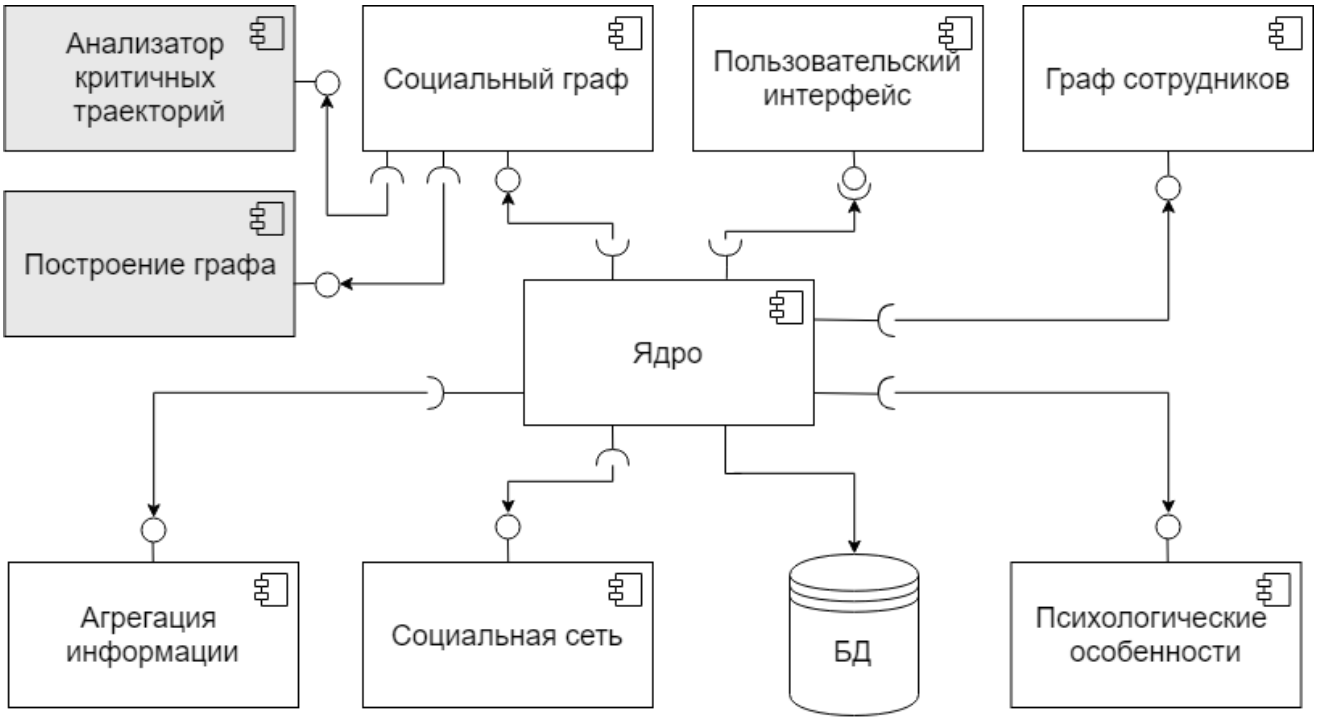


Рисунок 6 — Диаграмма программного комплекса

Разработанный программный модуль «Анализатор критичных траекторий» предназначен для реализации модели поиска наиболее вероятных и критичных траекторий распространения социоинженерных атак в социальном графе сотрудников компании. Модуль «Построение графа» осуществляет визуальное построение социального графа сотрудников компании, включая отображение наиболее критичных траекторий распространения, информацию о которых он получает из вышеописанного модуля. Более подробное описание данных модулей приводится в следующих двух разделах.

4.2. Выявление наиболее вероятной траектории распространения многоходовой социоинженерной атаки

Согласно подходу, предложенному в 3.2 поиск наиболее вероятной траектории распространения многоходовой социоинженерной атаки сводится к нахождению кратчайшего пути на социальном графе сотрудников компании. С учётом указанных в разделе 2.2 особенностей наиболее подходящими вариантами для решения задачи поиска наиболее вероятной траектории распространения являются алгоритмы Дейкстры и Беллмана–Форда. Данные алгоритмы позволяют обеспечить работу при ожидаемых вариациях исходных социальных графов сотрудников организаций. Для того чтобы добиться более быстрой работы алгоритма, без ущерба для точности, были введены дополнительные условия. А именно, для уменьшения вычислительной сложности алгоритма установлено пороговое значение: если оценка вероятности успешного распространения социоинженерной атаки от начальной вершины до обрабатываемой в данный момент вершины становится меньше заданного порога, то он исключается из рассмотрения.

Описанный подход был реализован в качестве модуля программного комплекса для автоматизированного анализа защищённости пользователей информационных систем от социоинженерных атак, схема и описание которого

представлены в [39]. Блок-схема алгоритма, заложенного в реализации данного модуля, представлена на Рисунок 7.

В качестве входных параметров используются идентификационные номера двух пользователей, наиболее критичную траекторию распространения социоинженерной атаки между которыми необходимо найти, а также социальный граф, получаемый с помощью одного из модулей указанного комплекса программ на основе данных, извлекаемых из контента, публикуемого пользователями в социальной сети ВКонтакте [43]. Результатом работы программного модуля является наиболее критичная траектория распространения социоинженерной атаки, оценка вероятности успеха прохождения по которой между двумя пользователями максимальна.

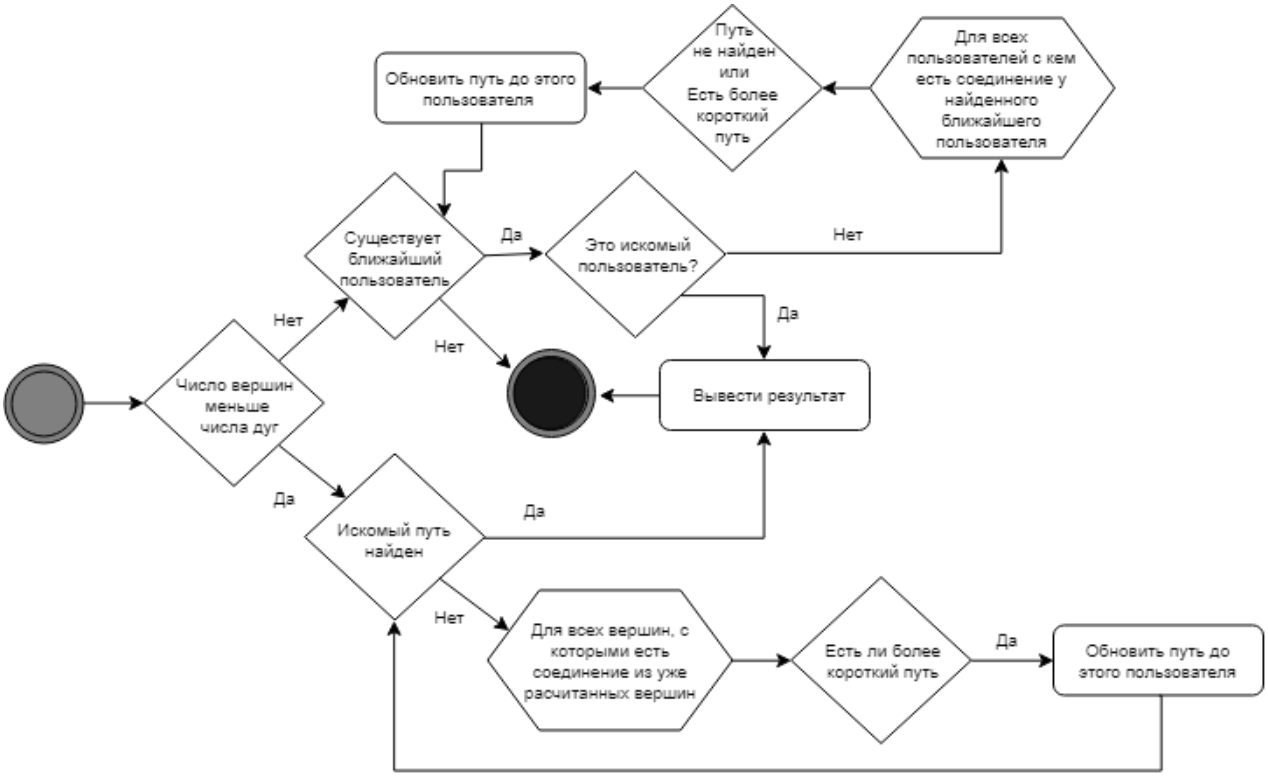


Рисунок 7 — Блок-схема алгоритма поиска наиболее вероятной траектории распространения социоинженерной атаки

Практическая значимость предложенного алгоритма заключается в расширении возможностей существующего программного комплекса и его дальнейшем применении при проверке безопасности пользователей информационных систем при многоходовых социоинженерных атаках

4.3. Визуализация социального графа сотрудников

Согласно проанализированным данным (Таблица 3) для наглядного представления социального графа сотрудников компании была выбрана библиотека GraphStream. Помимо достоинств, указанных в Таблица 3, ещё одним из преимуществ её использования является задание стилей графа имитирующее работу CSS.

Таблица 3 — Сравнение библиотек для представления социальных графов.

Название	Свободный доступ	Наглядность представления	Возможность изображения			Интерактивность	Итог
			Больших графов	Орграфов	Взвешенных графов		
JUNG	+	?	?	+	+	–	–
Neo4j	+	+	+	+	+	?	–
Graphviz	+	?	?	+	+	–	–
yFiles	–	+	+	+	+	?	–
Prefuse	+	?	?	+	+	+	–
JGraphT	+	–	–	+	+	–	–
GraphStream	+	+	+	+	+	+	+
JavaFX	+	–	–	+	+	+	–

На вход модуля по визуализации графа GraphBuilder подаётся социальный граф пользователей, принадлежащий классу SocialGraph. Подробная структура данного класса представлена в [38]. Сам граф может быть задан в программе, загружен из документа Microsoft Excel или получен из других программных модулей. Затем создаётся ориентированный мультиграф, узлы которого соответствуют структуре users, а рёбра — структуре connections. После чего узлам и рёбрам ставится в соответствие классы css: node и edge, с прописанными свойствами элементов. В связи с тем, что социальные графы пользователей обычно имеют большое число вершин и рёбер, была добавлена функция масштабирования графа, причём метки узлов и рёбер появляются только при определённом масштабе приближения. Также было реализовано представление

результатов работы модуля по нахождению наиболее вероятного пути: происходит выделение, входящих в него узлов и рёбер более тёмным цветом.

Таким образом, в ходе работы было реализовано: построение социального графа сотрудников на основе данных, выделение узла и рёбер по клику кнопки мыши, масштабирование графа, построение графа с выделением в нём наиболее вероятного пути. Кроме того, было осуществлено ориентированное представление социального графа. На рисунке 8 представлен пример работы программы, входными данными которой являлся документ Microsoft Excel с данными о сотрудниках компании со штатом 250 человек.

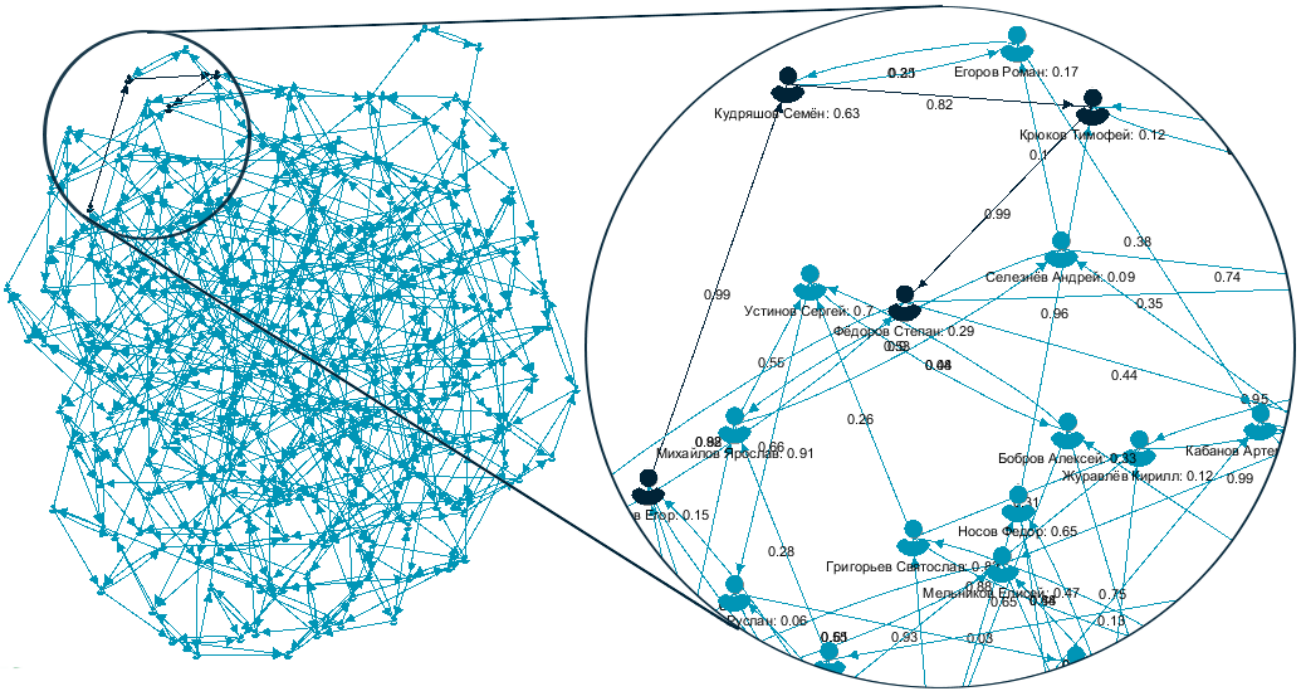


Рисунок 8 — Визуализация тестового графа сотрудников компании со штатом 250 человек.

Выводы по главе. В данной главе была приведена структура следующих программных модулей: «Анализатора критичных траекторий» и «Построения графа».

Заключение

Выпускная квалификационная работа бакалавра была посвящена разработке автоматизированных инструментов выявления наиболее критичных траекторий распространения многоходовых социоинженерных атак злоумышленника и их визуализации на социальном графе сотрудников. В рамках достижения данной цели были выполнены следующий задачи:

- предложен метод выявления наиболее вероятных траекторий распространения многоходовой социоинженерной атаки;
- предложена метрика для оценки наиболее критичных траекторий распространения многоходовой социоинженерной атаки;
- разработан и реализован алгоритм нахождения наиболее вероятных траекторий распространения многоходовых социоинженерных атак;
- предложены методы по изучению силы влияния возможных типов взаимоотношений между пользователями, на вероятность распространения социоинженерной атаки.

Обобщая вышеизложенное, все поставленный задачи были выполнены. Цель работы была успешно достигнута. Практическая значимость полученных результатов заключается в расширении возможностей существующего программного комплекса для анализа защищенности пользователей информационных систем от социоинженерных атак. Расширенный комплекс может быть рекомендован для использования в компаниях с целью проведения предупреждающей диагностики информационных сетей от социоинженерных атак. Перспективы дальнейшего исследования заключаются в рассмотрении моделей, которые более детально описывают контекст и учитывают распределение вероятностей поражения доли документов, доступных пользователю.

Список литературы

1. *Abawajy J. H., Ninggal M. I. H., Herawan T.* Privacy preserving social network data publication // IEEE communications surveys & tutorials. 2016. 18(3). pp. 1974 – 1997.

2. *Abramov M. V., Tulupyev A. L., Sulejmanov A. A.* Analysis of users' protection from socio-engineering attacks: social graph creation based on information from social network websites // Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2018. Vol. 18. № 2. pp. 313–321. doi: 10.17586/2226-1494-2018-18-2-313-321

3. *Albladi S. M., Weir G. R. S.* User characteristics that influence judgment of social engineering attacks in social networks. // Human-centric Computing and Information Sciences. 2018. 8(1). pp. 5

4. *Algarni A., Xu Y., Chan T.* An empirical study on the susceptibility to social engineering in social networking sites: the case of Facebook // European Journal of Information Systems. 2017. Vol. 26, № 6. pp. 661–687. doi: 10.1057/s41303-017-0057-y

5. *Azarov A. A., Tulupyeva T. V., Suvorova A. V., Tulupyev A. L., Abramov M.V., Usupov R.M.* Sotsioinzhenernye ataki: problemy analiza [Social engineering attacks: the problem of analysis] // St Petersburg: Nauka Publ., 2016. 349 p. (In Russian)

6. *Bhakta R., Harris I. G.* Semantic analysis of dialogs to detect social engineering attacks. // Semantic Computing (ICSC), 2015 IEEE International Conference on. – IEEE, 2015. pp. 424-427.

7. *Cai Z., He Z., Guan X., Li Y.* Collective data-sanitization for preventing sensitive information inference attacks in social networks // IEEE Transactions on Dependable and Secure Computing. 2018. № 15(4). pp. 577–590.

8. *Cao J., Fu Q., Li Q., Guo D.* Discovering hidden suspicious accounts in online social networks. // Information Sciences. 2017. № 394. pp. 123–140.

9. *Chiew K. L., Yong K. S. C., Tan C. L.* A survey of phishing attacks: their types, vectors and technical approaches. *Expert Systems with Applications*. 2018. doi:10.1016/j.eswa.2018.03.050

10. *Chin T., Xiong K., Hu C.* Phishlimiter: A Phishing Detection and Mitigation Approach Using Software-Defined Networking // IEEE Access. 2018. Vol. 6. pp. 42516–42531. doi:10.1109/ACCESS.2018.2837889

11. *Choi H. S., Lee W. S., Sohn S. Y.* Analyzing research trends in personal information privacy using topic modeling // Computers & Security 67. 2017. pp. 244–253.

12. *Cormen T. H., Leiserson C. E., Rivest R. L., Stein C.* Introduction to Algorithms // Second Edition MIT Press and McGraw-Hill. 2001. pp. 580–642.

13. *Curtis S. R., Rajivan P., Jones D. N., Gonzalez C.* Phishing attempts among the dark triad: Patterns of attack and vulnerability // Computers in Human Behavior. 2018. doi:10.1016/j.chb.2018.05.037

14. *Dang-Pham D., Pittayachawan S., Bruno V.* Why employees share information security advice? Exploring the contributing factors and structural patterns of security advice sharing in the workplace // Computers in Human Behavior 67. 2017. pp. 196 – 206.

15. *Dou Z., Khalil I., Khreishah A., Al-Fuqaha A., Guizani M.* Systematization of Knowledge (SoK): A Systematic Review of Software-Based Web Phishing Detection // IEEE Communications Surveys & Tutorials. 2017. Vol. 19. №. 4. pp. 2797–2819. doi:10.1109/COMST.2017.2752087

16. *Edwards M., Larson R., Green B., Rashid A., Baron A.* Panning for gold: automatically analysing online social engineering attack surfaces. // Computers & Security 69. 2017. pp. 18–34.

17. *Gupta B. B., Tewari A., Jain A. K., Agrawal D. P.* Fighting against phishing attacks: state of the art and future challenges // Neural Computing and Applications. 2017. Vol. 28, № 12. pp. 3629-3654. doi:10.1007/s00521-016-2275-y

18. *Jaafor O., Birregah B.* Multi-layered graph-based model for social engineering vulnerability assessment // Advances in Social Networks Analysis and Mining (ASONAM), 2015 IEEE/ACM International Conference on. IEEE. Springer, Paris. 2015. pp. 1480 – 1488.

19. *Junger M., Montoya L., Overink F. J.* Priming and warnings are not effective to prevent social engineering attacks. // Computers in human behavior. 2017. Vol. 66. pp. 75–87. doi:10.1016/j.chb.2016.09.012

20. *Kaur R., Singh S.* A comparative analysis of structural graph metrics to identify anomalies in online social networks. // Computers & Electrical Engineering 57. 2017. pp. 294 – 310.

21. *Lee K. C., Hsieh C. H., Wei L. J., Mao C. H., Dai J. H., Kuang Y. T.* Sec-Buzzer: cyber security emerging topic mining with open threat intelligence retrieval and timeline event annotation // Soft Computing. 2017. № 21(11). pp. 2883–2896.

22. *Levitin A.* Introduction to the design & analysis of algorithms // USA: Addison-Wesley. 2012. pp. 304–337

23. *Li H., Luo X. R., Zhang J., Sarathy R.* Self-control, organizational context, and rational choice in Internet abuses at work // Information & Management. 2018. № 55(3). pp. 358 – 367.

24. *Melville P., Mooney R., Nagarajan R.* Content-Boosted Collaborative Filtering for Improved Recommendations // University of Texas, USA: AAAI-02, Austin, TX, USA, 2002. pp. 187–192.

25. *Öğütçü G., Testik Ö. M., Chouseinoglou O.* Analysis of personal information security behavior and awareness // Computers & Security 56. 2016. pp. 83–93.

26. One Coffee? Your Total Is Some Personal Data. [Электронный ресурс] URL: http://nymag.com/selectall/2018/08/shiru-cafs-offer-students-free-coffee-for-harvested-data.html (Дата обращения: 27.09.2018)

27. Phishing campaign targets developers of Chrome extensions. [Электронный ресурс] URL: https://www.zdnet.com/article/phishing-campaign-targets-developers-of-chrome-extensions/ (Дата обращения: 08.10.2018)

28. Protecting People: A Quarterly Analysis of Highly Targeted Cyber Attacks. [Электронный ресурс] URL: https://www.proofpoint.com/us/resources/threat-reports/quarterly-threat-analysis (Дата обращения: 24.01.2018)

29. *Rassel S., Norvig P.* Artificial Intelligence: A Modern Approach // London: Prentice-Hall International. 2009. pp. 93-92

30. *Shindarev N., Bagretsov G., Abramov M., Tulupyeva T., Suvorova A.* Approach to identifying of employees profiles in websites of social networks aimed to analyze social engineering vulnerabilities // Advances in Intelligent Systems and Computing. Proceedings of the Second International Scientific Conference “Intelligent Information Technologies for Industry” (IITI’17). 2017. Vol. 1. pp. 441–447.

31. The White Company Series: Operation Shaheen Report [Электронный ресурс] URL: https://www.cylance.com/content/dam/cylance-web/en-us/resources/knowledge-center/resource-library/reports/WhiteCompanyOperationShaheenReport.pdf?_ga=2.171949088.1632831153.1542121197-1455929002.1542121197 (Дата обращения: 13.11.2018)

32. Warwick A. Social engineering is top hacking method, survey shows [Электронный ресурс] URL: https://www.computerweekly.com/news/4500272941/Social-engineering-is-top-hacking-method-survey-shows (дата обращения: 25.03.2018)

33. Wikipedia: Социальная инженерия [Электронный ресурс] URL: https://ru.wikipedia.org/wiki/Социальная_инженерия

34. Wikipedia: Социальная сеть [Электронный ресурс] URL: https://ru.wikipedia.org/wiki/Социальная_сеть

35. *Yang Z., Xue J., Yang X., Wang X., Dai Y.* VoteTrust: Leveraging friend invitation graph to defend against social network sybils // IEEE Transactions on Dependable and Secure Computing. 2016. № 13(4). pp. 488–501.

36. *Yasin A., Liu L., Li T., Wang J., Zowghi D.* Design and preliminary evaluation of a cyber Security Requirements Education Game (SREG) // Information and Software Technology. 2018. № 95. pp. 179–200.

37. *Zhang M., Qin S., Guo F.* Satisfying link perturbation and k-out anonymous in social network privacy protection // Communication Technology (ICCT), 2017 IEEE 17th International Conference on. – IEEE. IEEE Xplore, Chengdu. 2017. pp. 1387–1391.

38. *Абрамов М.В.* Методы и алгоритмы анализа защищенности пользователей информационных систем от социоинженерных атак: оценка параметров моделей: автореферат диссертации кандидата технических наук. Санкт-Петербургский институт информатики и автоматизации Российской академии наук, Санкт-Петербург, 2018. С. 148–154

39. *Абрамов М.В., Тулупьева Т.В., Тулупьев А.Л.* Социоинженерные атаки: социальные сети и оценки защищенности пользователей // СПб. ГУАП, 2018. 266 с. ISBN 978-5-8088-1377-5

40. Актуальные киберугрозы. I квартал 2018 года [Электронный ресурс] URL: https://www.ptsecurity.com/upload/corporate/ru-ru/analytics/Cybersecurity-threatscape-2018-Q1-rus.pdf (Дата обращения: 10.09.2018)

41. Атаки на информацию с помощью методов социальной инженерии [Электронный ресурс] URL: http://www.jetinfo.ru/stati/chelovek-cheloveku (дата обращения: 24.09.2018)

42. *Бычек В.* Социальная инженерия в интеллектуальной битве «добра» и «зла» [Электронный ресурс] *Защита информации.* Инсайд. 2006. № 6. С. 20–27.

43. Деньги с карт россиян киберворы стали снимать новым способом [Электронный ресурс] URL: http://www.amur.info/news/2018/09/05/143017 (Дата обращения: 02.10.2018)

44. *Каталков Д.* Как социальная инженерия открывает хакеру двери в вашу организацию // Positive Research 2018. Сборник исследований по практической безопасности. 2018. С. 26–30.

45. Киберпреступность в домашних тапочках [Электронный ресурс] URL: http://www.enforce.spb.ru/chronicle/publications-of-the-media/7130-aleksej-knorre-vedomosti-extra-jus-kiberprestupnost-v-domashnikh-tapochkakh

46. *Колесников Г. И., Хованов Н. В., Юдаева М. С.* Применение метода квантификации нечисловых оценок вероятности для выбора оптимального портфеля ценных бумаг // Вестник Санкт-Петербургского университета. Серия 5. Экономика. 2007. №. 3. С. 58–67.

47. На Avito замечена новая опасная схема развода россиян [Электронный ресурс] URL: https://zen.yandex.ru/media/apple-iphone.ru/na-avito-zamechena-novaia-opasnaia-shema-razvoda-rossiian-5bffe7ad64b5010fabb426c1 (Дата обращения: 22.12.2018)

48. Российская Федерация. Законы. Закон Российской Федерации «Об информации, информатизации и защите информации». Федер. закон: в ред. от 10.01.2003 №15-ФЗ. М. Ось-89, 2005. 32 с.

49. Сбербанк: Основной угрозой для клиентов является социальная инженерия [Электронный ресурс] URL: https://www.anti-malware.ru/news/2017-12-07-1447/25019 (дата обращения: 15.03.2018)

50. Специальные требования и рекомендации по технической защите конфиденциальной информации (СТР-К). Решение Коллегии Гостехкомиссии России №7/02.03.01 г. [Электронный ресурс] URL: http://www.confidentiality.strongdisk.ru

51. *Сулейманов А.А., Абрамов М.В.* Автоматизация построения социального графа сотрудников компании на основе публикуемого ими контента в социальных сетях // Школа-семинар по искусственному интеллекту: сборник научных трудов. Тверь: ТвГТУ. 2018. С. 32–40.

52. *Сулейманов А.А., Абрамов М.В., Тулупьев А.Л.* Оценка вероятности поражения критичного документа при многоходовых социоинженерных атаках //

Сборник докладов Международной конференции по мягким вычислениям и измерениям (SCM-2018). Санкт-Петербург. Том 1-2. Т. 1. 2018. С. 130–133.

53. ФНС России предупреждает о создании мошеннического сайта-клона [Электронный ресурс] URL: https://www.nalog.ru/rn77/news/activities_fts/7245895/ (дата обращения: 14.08.2018)

54. Хакеры украли у россиян миллионы рублей через лжесайты [Электронный ресурс] URL: https://news.ru/den-gi/hakery-pohitili-u-rossiyan-svyshe-250-mln-rublej-cherez-lzhesajty/?bulk_email_rid=259&contactId=c71d6b02-74c9-4a59-8227-44fe0265622e&bulkEmailRecipientId=94c550cc-95c4-4a5b-be33-4ed062e77878/

55. *Хованов Н. В., Федотов Ю. В.* Модели учета неопределённости при построении сводных показателей эффективности деятельности сложных производственных систем // Научные доклады. 2006. №. 28R-2006. 37 с.

56. ЦБ ожидает роста активности мошенников, использующих социальную инженерию. [Электронный ресурс] URL: https://ria.ru/economy/20171213/1510861611.html (дата обращения: 07.05.2018)

Список иллюстративного материала

Список рисунков:

Список таблиц:

Приложение А: словарь терминов

Безопасность информации — состояние защищенности информации, характеризуемое способностью персонала, технических средств и информационных технологий обеспечивать конфиденциальность (т.е. сохранение в тайне от субъектов, не имеющих полномочий на ознакомление с ней), целостность и доступность информации при ее обработке техническими средствами [49].

Социальный граф пользователей — граф, узлы которого представлены социальными объектами, такими как пользовательские профили с различными атрибутами (например: имя, день рождения, родной город и т.д.), сообщества, медиа-контент и т.д., а ребра — социальными связями между ними [24].

Документ — материальный объект с зафиксированной на нем информацией в виде текста, звукозаписи или изображения, предназначенный для передачи во времени и пространстве в целях хранения и общественного использования [48].

Злоумышленник — социальный инженер, целью которого является получение доступа к конкретному документу/ряду документов при помощи методов социальной инженерии, зачастую эксплуатируя доверчивость, лень, любезность и энтузиазм пользователей и сотрудников организаций. [33]

Многоходовая социоинженерная атака — социоинженерная атака, которая включает в себя взлом более чем одного сотрудника таким образом, что взломанные сотрудники непосредственно участвуют во взломе последующих жертв. [2]

Профиль уязвимостей пользователя — совокупность пар «уязвимость пользователя» - «степень выраженности уязвимости» [38].

Социальная сеть — платформа, онлайн-сервис и веб-сайт, предназначенные для построения, отражения и организации социальных взаимоотношений в Интернете [34].

Социоинженерная атака — набор прикладных психологических и аналитических приёмов, которые злоумышленник применяет для скрытой мотивации пользователей публичной или корпоративной сети к нарушениям устоявшихся правил и политик в области информационной безопасности [38].

Уязвимость пользователя — некоторая характеристика пользователя которая делает возможным успешное завершение социоинженерного атакующего действия злоумышленника [38].

Приложение Б: перечень публикаций

Научные тезисы и статьи, опубликованные по теме выпускной квалификационной работы бакалавра:

1. **Азаров А. А., Хлобыстова А. О.** Построение стратегии защиты пользователей информационных систем от социоинженерных атакующих воздействий злоумышленника. // *Информационная безопасность регионов России (ИБРР-2017). Юбилейная X Санкт-Петербургская межрегиональная конференция.* Санкт-Петербург, 1-3 ноября *2017* с. 410-411

2. **Хлобыстова А. О., Абрамов М.В., Тулупьев А.Л.** Идентификация наиболее вероятных траекторий социоинженерных атак в управлении рисками, ассоциированными с пользователями/персоналом // *Сборник материалов конференции «Информационные технологии в управлении» (ИТУ-2018),* Санкт-Петербург, 2018. с. 493–497

3. **Хлобыстова А.О., Абрамов М.В.** Выявление наиболее критичных траекторий распространения многоходовых социоинженерных атак // *Региональная информатика (РИ-2018). XVI Санкт-Петербургская международная конференция «Региональная информатика (РИ-2018)».* (Санкт-Петербург, 24-26 октября 2018 г.): Материалы конференции. СПб: СПОИСУ, 2018. С. 561–562

4. **Khlobystova A.O., Abramov M.V., Tulupyev A.L.** Identifying the most critical trajectory of the spread of a social engineering attack between two users // *Fuzzy Technologies in the Industry (FTI 2018),* Ulyanovsk, 2018. Pp. 38–43

5. **Хлобыстова А. О., Абрамов М. В., Тулупьев А. Л., Золотин А. А.** Выявление наиболее критичной траектории распространения социоинженерной атаки между двумя пользователями. // *Информационно-управляющие системы*, 2018, № 6, с. 33– 40. doi:10.31799/1684-8853-2018-6-74-81

6. **Khlobystova A.O., Abramov M.V., Tulupyev A.L.** An approach to estimating of criticality of social engineering attacks traces // Studies in Systems, Decision and Control. P. 446–456. ISSN: 2198-4182

Работа выполнена при финансовой поддержке РФФИ, проект №18-37-00340 – Методы анализа устойчивости структуры социальных связей пользователей информационной системы к социоинженерным атакующим воздействиям злоумышленника на основе применения генетических алгоритмов.

По разработанным в ходе исследовательской работы модулям были получены два свидетельства о регистрации программы в Роспатенте:

1. **Хлобыстова А.О., Абрамов М.В., Тулупьев А.Л.** Identifying critical trajectory of the spread of a social engineering attack, Version 01 for Java (ICTS SEA jav.v.01) (Свидетельство). Свид. о гос. рег. прогр. для ЭВМ. Рег. № 2019610872 (12.02.2019)

2. **Хлобыстова А.О., Абрамов М.В., Тулупьев А.Л, Тулупьева Т.В.** Quantification of relationship represented in the social network, Version 01 for CSharp (QR SN cs.v.01) (Свидетельство). Свид. о гос. рег. прогр. для ЭВМ. Рег. № 2019610870 (12.02.2019)